IN THE CLAIMS:

Please substitute the following claims for the same-numbered claims in the application:

1. (Currently Amended) A method for text summarization produced by clustering data points with defined quantified relation values between them, said method comprising:

obtaining a lead value for each data point, wherein said lead value for each data point is ~~derived~~ calculated by taking a sum of all relation values input into said data point ~~plus~~ weighted by a frequency of occurrence associated with said data point,

ranking each data point in a lead value sequence list in descending order of lead value,

assigning a first data point in said lead value sequence list as a leader of a first cluster,

considering each subsequent data point in said lead value sequence list as a leader of a new cluster if its relationship with leaders of each of the previous clusters is less than a defined threshold value or as a member of at least one cluster where its relationship with a cluster leader is at least equal to said threshold value, wherein the threshold value is adaptively found for a given number of clusters, and

generating [[a]] said text summarization of any of a single document and a collection of documents by segmenting a given text input comprising said data points into clusters, and forming a set of leaders of said clusters to represent said text summarization.

2. (Previously Presented) The method of claim 1, wherein said quantified relationships between data points are any of symmetric and asymmetric quantified relationships.

3. (Currently Amended) The method of claim 1, wherein said frequency of occurrence equals one.

4. (Previously Presented) The method of claim 1, further comprising identifying distinct data points using said lead values and said relation values between said data points.

5. (Previously Presented) The method of claim 1, further comprising organizing a set of data points into a hierarchy of clusters by clustering the data points into sets of small sizes, wherein each smaller set is further subclustered; and repeatedly subclustering said smaller set until a terminating condition is reached.

6. (Previously Presented) The method of claim 1, wherein said step of generating further comprises:

segmenting a given input text into blocks comprising sentences, a collection of sentences, and paragraphs,

excluding words belonging to a defined list of defined stop words,

replacing words by their existing unique synonymous word from a given a collection of synonyms,

applying stemming algorithms for mapping words to root words,

representing resulting blocks of text, with respect to a dictionary which is either given or computed from the input text, by a binary vector of size equal to the number of words in the dictionary whose ith element is 1 if an ith word in the dictionary is present in the block,

computing the relationship between any data points di and dj by evaluating $R(di,dj) = |di.Tdi|/|di|$, wherein T is a thesaurus matrix whose ijth element reflects an extent of inclusion of meaning of jth word in the meaning of ith word, and

clustering the data points.

7. (Previously Presented) The method of claim 6, wherein said dictionary is computed by taking a fraction of words, excluding said stop words, with a highest tfidf value, which is given by:

tfidf(wi) = tfi * log(N/dfi),

where tfidf(wi) is the lead value of data point wi, tfi – a number of times the data point wi occurred in a whole text, dfi = a number of documents containing the data point wi and N = the a total number of documents in the text.

8. (Previously Presented) The method of claim 6, wherein said thesaurus matrix comprises any of a given identity matrix, and a computed matrix from a collection of documents.

9. (Currently Amended) The method of claim 6, wherein each block is represented by a vector whose ith element represents [[a]] said frequency of occurrence of said ith word in the block.

10. (Previously Presented) The method of claim 6, further comprising organizing a set of text documents into a hierarchy of clusters by clustering given documents into sets of small sizes, wherein each smaller set is further subclustered; and repeatedly subclustering said smaller set until a terminating condition is reached.

11. (Previously Presented) The method of claim 10, further comprising organizing results

returned by an information retrieval system in response to an user query into an hierarchy of clusters.

12. (Previously Presented) The method of claim 11, wherein the hierarchy is used to aid the user in any of modifying a query of said user and browsing through said results.

13. (Previously Presented) The method of claim 11, wherein said information retrieval system comprises a search engine retrieving Web documents.

14. (Previously Presented) The method of claim 5, wherein said step of generating is applied to vocabulary organization for a group of documents wherein the data points are words in a dictionary of the vocabulary, wherein the lead value of a word is any of its frequency of occurrence in the collection of documents, a number of documents containing the word, and a tfidf value of said word, wherein a relationship R(di,dj) denotes a fraction of documents containing a jth word that also contains an ith word, and the clustering of said data points results in a structured hierarchical organization of the vocabulary.

15. (Previously Presented) The method of claim 14, wherein a structured vocabulary is used to provide text summarization for associated documents.

16. (Previously Presented) The method of claim 14, further comprising applying the clustering to customer profiling wherein a dictionary is built and the vocabulary is organized using documents that are viewed by a customer.

17. (Previously Presented) The method of claim 5, wherein said data points correspond to products cataloged in an electronic store, the lead value of a product is its per unit profit, its per unit value or a number of items sold per unit time, and a relationship between the products is either explicitly defined or derived from purchase data.

18. (Previously Presented) The method of claim 17, wherein a product di is related to a product dj by the a fraction of customer transactions containing dj that also contain di.

19. (Previously Presented) The method of claim 17, further comprising applying the clustering to any of to an analysis of sales of a store for a merchant, and an organization of a layout of the store to facilitate easy access to products.

20. (Previously Presented) The method of claim 17, further comprising applying the clustering to personalize an electronic store layout to an individual customer by using a relationship that is specific to the individual customer.

21. (Previously Presented) The method of claim 5, further comprising applying the clustering to customer segmentation for a sales or service organization wherein the data points comprise customers in a database, wherein the lead values are any of a total purchase amount per unit time of said customers, income of said customers, a number of times customers visited an electronic store, and a number of items bought by the customer, wherein a relationship between customers is either explicitly defined or derived from some relevant data, with a resulting clustering

reflecting a structured grouping of customers with similar performances.

22. (Previously Presented) The method of claim 21, wherein a customer di is related to a customer dj by a fraction of products bought by dj that are also bought by di.

23. (Currently Amended) A system for text summarization produced by clustering data points with defined quantified relation values between them, said system comprising:

means for obtaining a lead value for each data point, wherein said lead value for each data point is ~~derived~~ calculated by taking a sum of all relation values input into said data point ~~plus~~ weighted by a frequency of occurrence associated with said data point,

means for ranking each data point in a lead value sequence list in descending order of lead value,

means for assigning a first data point in said lead value sequence list as a leader of a first cluster,

means for considering each subsequent data point in said lead value sequence list as a leader of a new cluster if its relationship with leaders of each of the previous clusters is less than a defined threshold value or as a member of at least one cluster where its relationship with a cluster leader is at least equal to said threshold value, wherein the threshold value is adaptively found for a given number of clusters, and

means for generating [[a]] said text summarization of any of a single document and a collection of documents by segmenting a given text input comprising said data points into clusters, and forming a set of leaders of said clusters to represent said text summarization.

24. (Previously Presented) The system of claim 23, wherein said quantified relationships between data points are any of symmetric and asymmetric quantified relationships.

25. (Currently Amended) The system of claim 23, wherein said frequency of occurrence equals one.

26. (Previously Presented) The system of claim 23, further comprising means for identifying distinct data points using said lead values and said relation values between said data points.

27. (Previously Presented) The system of claim 23, further comprising means for organizing a set of data points into a hierarchy of clusters using means for clustering the data points into sets of small sizes, wherein each smaller set is further subclustered; and repeatedly subclustering said smaller set until a terminating condition is reached.

28. (Previously Presented) The system of claim 23, wherein said means for generating further comprises:

means for segmenting a given input text into blocks comprising sentences, a collection of sentences, and paragraphs,

means for excluding words belonging to a defined list of defined stop words,

means for replacing words by their existing unique synonymous word from a given a collection of synonyms,

means for applying stemming algorithms for mapping words to root words,

means for representing resulting blocks of text, with respect to a dictionary which is

either given or computed from the input text, by a binary vector of size equal to the number of words in the dictionary whose ith element is 1 if an ith word in the dictionary is present in the block,

means for computing the relationship between any data points di and dj by evaluating R(di,dj) = |di.Tdi|/|di|, wherein T is a thesaurus matrix whose ijth element reflects an extent of inclusion of meaning of jth word in the meaning of ith word, and

means for clustering the data points.

29. (Previously Presented) The system of claim 28, wherein said dictionary is computed by taking a fraction of words, excluding said stop words, with a highest tfidf value, which is given by:

tfidf(wi) = tfi * log(N/dfi),

where tfidf(wi) is the load value of data point wi, tfi = a number of times the data point wi occurred in a whole text, dfi = a number of documents containing the data point wi and N = a total number of documents in the text.

30. (Previously Presented) The system of claim 28, wherein said thesaurus matrix comprises any of a given identity matrix, and a computed matrix from a collection of documents.

31. (Currently Amended) The system of claim 28, wherein each block is represented by a vector whose ith element represents [[a]] said frequency of occurrence of said ith word in the block.

32. (Previously Presented) The system of claim 28, further comprising means for organizing a set of text documents into a hierarchy of clusters by using means for clustering given documents into sets of small sizes, wherein each smaller set is further subclustered; and repeatedly subclustering said smaller set until a terminating condition is reached.

33. (Previously Presented) The system of claim 32, further comprising means for organizing results returned by an information retrieval system in response to an user query into an hierarchy of clusters.

34. (Previously Presented) The system of claim 33, wherein the hierarchy is used to aid the user in any of modifying a query of said user and browsing through said results.

35. (Previously Presented) The system of claim 33, wherein said information retrieval system comprises a search engine retrieving Web documents.

36. (Previously Presented) The system of claim 27, wherein said means for generating is used for vocabulary organization for a group of documents wherein the data points are words in a dictionary of the vocabulary, wherein the lead value of a word is any of its frequency of occurrence in the collection of documents, a number of documents containing the word, and a tfidf value of said word, wherein a relationship R(di,dj) denotes the a fraction of documents containing a jth word that also contains an ith word, and the clustering of said data points results in a structured hierarchical organization of the vocabulary.

37. (Previously Presented) The system of claim 36, wherein a structured vocabulary is used to provide text summarization for associated documents.

38. (Previously Presented) The system of claim 36, further comprising means for using the clustering for customer profiling wherein a dictionary is built and the vocabulary is organized using documents that are viewed by a customer.

39. (Previously Presented) The system of claim 27, wherein said data points correspond to products cataloged in an electronic store, the lead value of a product is its per unit profit, its per unit value or a number of items sold per unit time, and a relationship between the products is either explicitly defined or derived from purchase data.

40. (Previously Presented) The system of claim 39, wherein a product di is related to a product dj by a fraction of customer transactions containing dj that also contain di.

41. (Previously Presented) The system of claim 39, further comprising means for applying the clustering to any of an analysis of sales of a store for a merchant, and an organization of a layout of the store to facilitate easy access to products.

42. (Previously Presented) The system of claim 39, further comprising means for applying the clustering to personalize an electronic store layout to an individual customer by using a relationship that is specific to the individual customer.

43. (Previously Presented) The system of claim 27, further comprising means for applying the clustering for customer segmentation for a sales or service organization wherein the data points comprise customers in a database, wherein the lead values are any of a total purchase amount per unit time of said customers, income of said customers, a number of times customers visited an electronic store, and a number of items bought by the customer, wherein a relationship between customers is either explicitly defined or derived from some relevant data, with a resulting clustering reflecting a structured grouping of customers with similar performances.

44. (Previously Presented) The system of claim 43, wherein a customer di is related to a customer dj by a fraction of products bought by dj that are also bought by di.

45. (Currently Amended) A computer program product comprising computer readable program code stored on computer readable storage medium embodied therein for text summarization produced by clustering data points with defined quantified relation values between them, said computer program product comprising:

computer readable program code means for obtaining a lead value for each data point, wherein said lead value for each data point is ~~derived~~ calculated by taking a sum of all relation values input into said data point ~~plus~~ weighted by a frequency of occurrence associated with said data point,

computer readable program code means for ranking each data point in a lead value sequence list in descending order of lead value,

computer readable program code means for assigning a first data point in said lead value sequence list as a leader of a first cluster,

computer readable program code means for considering each subsequent data point in said lead value sequence list as a leader of a new cluster if its relationship with leaders of each of the previous clusters is less than a defined threshold value or as a member of at least one cluster where its relationship with a cluster leader is at least equal to said threshold value, wherein the threshold value is adaptively found for a given number of clusters, and

computer readable program code means for generating [[a]] said text summarization of any of a single document and a collection of documents by segmenting a given text input comprising said data points into clusters, and forming a set of leaders of said clusters to represent said text summarization.

46. (Previously Presented) The computer program product of claim 45, wherein said quantified relationships between data points are any of symmetric and asymmetric quantified relationships.

47. (Currently Amended) The computer program product of claim 45, wherein said frequency of occurrence equals one.

48. (Previously Presented) The computer program product of claim 45, further comprising computer readable program code means for identifying distinct data points using said lead values and said relation values between said data points.

49. (Previously Presented) The computer program product of claim 45, further comprising computer readable program code means configured for organizing a set of data points into a

hierarchy of clusters using computer readable program code means configured for clustering the data points into sets of small sizes, wherein each smaller set is further subclustered; and repeatedly subclustering said smaller set until a terminating condition is reached.

50. (Previously Presented) The computer program product of claim 45, wherein said computer readable program code means configured for generating further comprises:

computer readable program code means configured for segmenting a given input text into blocks comprising sentences, a collection of sentences, and paragraphs,

computer readable program code means configured for excluding words belonging to a defined list of defined stop words,

computer readable program code means configured for replacing words by their existing unique synonymous word, if it exists, from a given a collection of synonyms,

computer readable program code means configured for applying stemming algorithms for mapping words to root words,

computer readable program code means configured for representing resulting blocks of text, with respect to a dictionary which is either given or computed from the input text, by a binary vector of size equal to the number of words in the dictionary whose ith element is 1 if an ith word in the dictionary is present in the block,

computer readable program code means configured for computing the relationship between any data points di and dj by evaluating $R(di,dj) = |di.Tdi|/|di|$, wherein T is a thesaurus matrix whose ijth element reflects an extent of inclusion of meaning of jth word in the meaning of ith word, and

computer readable program code means configured for clustering the data points.

51. (Previously Presented) The computer program product of claim 50, wherein said dictionary is computed by taking a fraction of words, excluding said stop words, with a highest tfidf value, which is given by:

tfidf(wi) = tfi * log(N/dfi),

where tfidf(wi) is the lead value of data point wi, tfi = a number of times the data point wi occurred in a whole text, dfi = a number of documents containing the data point wi and N = a total number of documents in the text.

52. (Previously Presented) The computer program product of claim 50, wherein said thesaurus matrix comprises any of a given identity matrix, and a computed matrix from a collection of documents.

53. (Currently Amended) The computer program product of claim 50, wherein each block is represented by a vector whose ith element represents [[a]] said frequency of occurrence of said ith word in the block.

54. (Previously Presented) The computer program product of claim 50, further comprising computer readable program code means configured for organizing a set of text documents into a hierarchy of clusters by using computer readable program code means configured for clustering given documents into sets of small sizes, wherein each smaller set is further subclustered; and repeatedly subclustering said smaller set until a terminating condition is reached.

55. (Previously Presented) The computer program product of claim 54, further comprising computer readable program code means configured for organizing results returned by an information retrieval system in response to an user query into an hierarchy of clusters.

56. (Previously Presented) The computer program product of claim 55, wherein the hierarchy is used to aid the user in any of modifying a query of said user and browsing through said results.

57. (Previously Presented) The computer program product of claim 55, wherein said information retrieval system comprises a search engine retrieving Web documents.

58. (Previously Presented) The computer program product of claim 49, wherein said computer readable program code means configured for generating is used for vocabulary organization for a group of documents wherein the data points are words in a dictionary of the vocabulary, wherein the lead value of a word is any of its frequency of occurrence in the collection of documents, a number of documents containing the word, and a tfidf value of said word, wherein a relationship R(di,dj) denotes a fraction of documents containing a jth word that also contains an ith word, and the clustering of said data points results in a structured hierarchical organization of the vocabulary.

59. (Previously Presented) The computer program of claim 58, wherein a structured vocabulary is used to provide text summarization for associated documents.

60. (Previously Presented) The computer program product of claim 58, further comprising

computer readable program code means configured for using the clustering for customer profiling wherein a dictionary is built and the vocabulary is organized using documents that are viewed by a customer.

61. (Previously Presented) The computer program product of claim 49, wherein said data points correspond to products cataloged in an electronic store, the lead value of a product is its per unit profit, its per unit value or a number of items sold per unit time, and a relationship between the products is either explicitly defined or derived from purchase data.

62. (Previously Presented) The computer program product of claim 61, wherein a product di is related to a product dj by a fraction of customer transactions containing dj that also contain di.

63. (Previously Presented) The computer program product of claim 61, further comprising computer readable program code means configured for applying the clustering to any of an analysis of sales of a store for a merchant, and an organization of a layout of the store to facilitate easy access to products.

64. (Previously Presented) The computer program product of claim 61, further comprising computer readable program code means configured for applying the clustering to personalize an electronic store layout to an individual customer by using a relationship that is specific to the individual customer.

65. (Previously Presented) The computer program product of claim 49, further comprising

computer readable program code means configured for applying the clustering for customer segmentation for a sales or service organization wherein the data points comprise customers in a database, wherein the lead values are any of a total purchase amount per unit time of said customers, income of said customers, a number of times customers visited an electronic store, and a number of items bought by the customer, wherein a relationship between customers is either explicitly defined or derived from some relevant data, with a resulting clustering reflecting a structured grouping of customers with similar performances.

66. (Previously Presented) The computer program product of claim 65, wherein a customer di is related to a customer dj by a fraction of products bought by dj that are also bought by di.